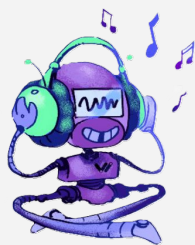


# Stenograf

Exploration of training Zipformer and E-Branchformer models with Polish language BIGOS dataset.



# Nasz Wkład

**Pomagaliśmy stworzyć zbiór testowy przez korektę manualną**  
dlatego podjęliśmy decyzję, że nie zgłosimy naszego modelu

# Nasze zgłoszenie

**102 %** to było zgłoszenie testowe.

test-A CER score	test-A WER score	test-B CER score	test-B WER score
0.9397742349435587	1.0180248042829088	0.9472076982485419	1.0206505898204417

Chciałem mieć błąd bliski 100% dlatego dokonałem manipulacji wyniku z modelu ASR tak by zamienić każdą literę w słowie.

# Zipformer

## Zoptymalizowana wydajna wersja Transformera, arch. Encoder-decoder

U-Net-like encoder structure where middle stacks operate at lower frame rates;

multi-scale attention, łączy lokalne i globalne zależności

sparsity-based attention i inne optymalizacyjne

Swoosh activation functions - modified LayerNorm

inputs: 80-dimension Mel filter-bank (25ms, hop 10ms)

model : size M (65.6M params)

optimizer: ScaledAdam

hyper params: default from k2 Icefall framework

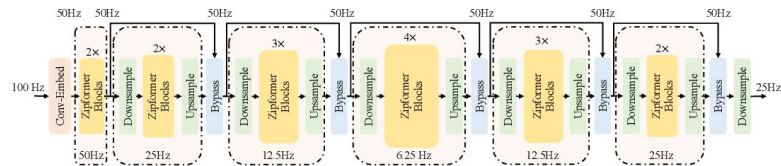
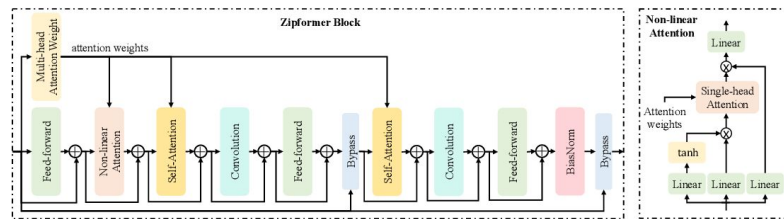


Figure 1: Overall architecture of Zipformer.



# E-Branchformer

## E-Branchformer - typ conformer

dzieli proces przetwarzania na kilka równoległych "ścieżek"  
które następnie integrują informacje  
w sposób bardziej elastyczny i wydajny.  
dynamic attention

inputs: 80-im log Mel filter-bank (32ms, hop 10ms)

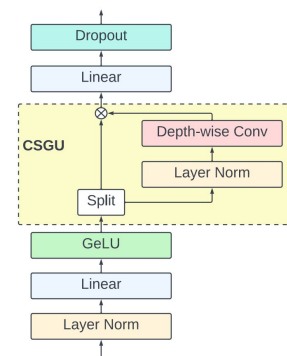
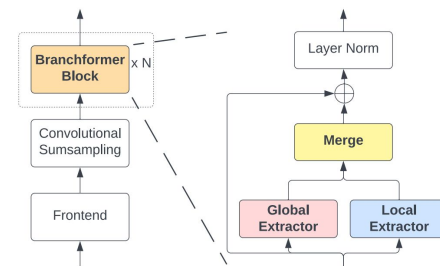
output: 5K BPE sub-word units as output tokens

model : size M (53.71M params)

16 layers, 128 embed dim., 512 attention dim. and 8 attention heads

optimizer: ScaledAdam

hyper params: default from ESPnet framework



: A figure of the Local extractor branch in Branchformer.

# Wynik

Table 1: Results of fined-tuned models.

	<b>dev-0 WER</b>
Zipformer	21.52
E-Branchformer	19.23

Pytania ?  
questions ?



Contact me:  
twitter: [@cyrta](#)  
or  
[pawel@cyrta.com](mailto:pawel@cyrta.com)